AD-A117 112 MICHIGAN UNIV ANN ARBOR SYSTEMS ENGINEERING LAB F/G 9/2

VECTORIZED SPARSE ELIMINATION.(U)

MAY 82 D A CALAHAN AFOSR-80-0158

UNCLASSIFIED AFOSR-TR-82-0553 NL

1 OF 1

END
DATE
FILMED
8-82
DTIC

(4)

AFOSR-TR- 82-0553

AD A117112

Interim Technical Report
for Period 5-1-81 to 4-30-82
Grant -AFOSR-80-0158

VECTORIZED SPARSE ELIMINATION

D. A. Calahan
Principal Investigator
Systems Engineering Laboratory
University of Michigan
Ann Arbor, MI

April 15, 1982

82 07 19 040

A. *INTRODUCTION*

Figure 1 indicates how the single topic of general sparse matrix solution using scalar processors may be broken into specialized areas of study when implementation on vector architectures is considered.

First, highly sparse matrices, usually representing ODE/algebraic-modeled systems, are easily decoupled by re-ordering. At a minimum, locally-decoupled equations may be solved in pipelined scalar mode (see below); if the decoupled subsystems can be arranged (a) to have identical sparsity, and (b) to be stored a constant stride apart, then a simultaneous sparse solver [7] may be invoked and a vector solution obtained.

As sparse systems become locally coupled - as occurs in finite element and finite difference problems - then vectors are easily defined within the coupled subsystems. It is worth making a further distinction between

> (a) intra-nodal or intra-element coupling, where the dimension of dense submatrices is proportional to the number of unknowns/node or unknowns/finite element, and
>
> (b) inter-nodal or inter-element, where the coupling between grid nodes or finite elements determines the vector length.

Banded and profile matrices result from the latter. The associated vector lengths are the *products* of the number of unknown/node (element) and the number of coupled nodes. These lengths are therefore always longer than in the former case, so that common bandsolvers offer the highest performance of any sparse solvers.

In previous research, algorithms and CRAY-1 software have been developed on this grant for (Figure 1)

> (a) general sparse matrices [10],
>
> (b) patterned sparse matrices, in conjunction with a vectorized electronic

circuit analysis program [7][11], and

(c) blocked matrices arising from intra-nodal coupling [8].

In addition, Duff and Reid [9] have converted a Fortran frontal ("intra-element") solver to the CRAY-1.

B. PROGRESS

1. SPARSE MATRICES

Referring to Figure 1 again, (a)-(c) above leave

(d) unpatterned highly sparse matrices, and

(e) banded and profile matrices

for current study.

UNPATTERNED

A highly- and randomly-sparse matrix necessitates a scalar solution. Even so, an ordering (in the spirit of nested dissection) can be found such that the LU factors can be written in the form

$$\begin{array}{cccc} D_{11} & & U_{12} & \\ & D_{22} & & U_{23} \\ L_{21} & & D_{33} & \\ & L_{32} & & \end{array}$$

where D_{ii} is a diagonal matrix and L_{ij} and U_{ji} extend from D_{ii} to the matrix boundary. The ordered steps to reduce the rth pivot block are

$$D_{rr} \;\leftarrow\; D_{rr}^{-1} \qquad \text{(reciprocation)} \qquad (1)$$

$$U_{r,r+1} \;\leftarrow\; D_{rr}A_{r,r+1} \qquad \text{(multiplication)} \qquad (2)$$

$$A_{r+1,r+1} \;\leftarrow\; A_{r+1,r+1}L_{r+1,r}U_{r,r+1} \qquad \text{(multi./subtraction)} \qquad (3)$$

where $A_{r+1,r+1}$ represents the unreduced southeast corner of the matrix at the rth reduction step. These steps can be performed in three parallel or pipelined steps.

Because all of the above operations are on blocks with random sparsity and storage, they must be performed in scalar mode. We have elected to achieve the highest speed by generating loopless scalar solution code in the manner of Gustavson [12], although this limits the matrix size to perhaps 5,000 highly-sparse equations. Rates in the range of 15 MFLOPS on the CRAY-1 are readily achievable, a speedup of 5:1 to 10:1 over other implementations. A report and a paper are currently in preparation on these equation ordering and implementation methods.

BANDED/PROFILE

It has become popular among researchers recently studying the solution of 2-D grids by general sparse solvers on the CRAY-1 to quibble about the quality of software operating in the range of 5-20 MFLOPS. Yet it is well known that banded solution of such grids is possible in the range of 90 MFLOPS [13]. It was therefore deemed desirable to develop a high-performance bandsolver as a standard for comparison with lower-performance general sparsity software. General sparsity software would then be useful only if the reduced operation counts associated with optimal ordering could compensate for the higher performance of a banded solver. This bandsolver was completed in the Fall of 1981. Careful CAL coding using a CRAY-1 simulator resulted in a 1.3:1 to 2:1 speedups over the best previous (Los Alamos) coding, to over 117 MFLOPS.

It is well-known that factors of up to four in operation count may be achieved in banded solutions of common irregular grid structures by following the bandwidth profile. Unfortunately, the pointers to describe this profile constitute a serious overhead in the inner loops of the vectorized reduction algorithm. It was therefore decided to *block* the profile parallel to the diagonal (Figure 2). Diagonal blocking is natural to 2-D grid solution, and so introduces few extra computations; it also migrates the symbolic pointers to the outer loops, since block descriptors point at large matrix substructures. Overall, this appears to be the best compromise between the operation count efficiency of general sparsity methods and the vectorizability of banded solution.

Software has been developed for the CRAY-1 to solved banded and profile systems [2][5] and has been included in a library of sparse equation solvers classified in Figure 1 and directed at general and specialized sparsity structures. (These solvers will be presented at a forthcoming sparse matrix conference in Fall, 1982.) A symmetric matrix version of this software is intended to be developed with joint AFFDL support; it will be used to solve optimization problems in the structural aspects of wing design. Also, optimal blocking strategies are being studied, based on a timing model (developed using a CRAY-1 simulator) of the solution code.

2. OTHER PROGRESS

SCHEDULING

It was observed during the development of high performance equation solvers for the CRAY-1 that optimality of the implementation could not be guaranteed. It appeared, however, that the CRAY-1 architecture could be described as a mathematical programming problem. Consistent with our investment in other program development aids, this optimizer is being developed into a useful package and is apparently of interest to others with desire for truly optimal codings. Moreover, a related conference presentation has been made [3], a journal manuscript is being prepared, and a Ph.D. thesis is being written on this work.

ELECTRONIC CIRCUIT ANALYSIS

Because our multi-level sparse matrix algorithms are the basis of the Berkeley effort to vectorize their popular SPICE electronic circuit analysis program, it is worth reporting that their AFOSR-funded project is producing speedups of 6:1 over the original scalar code for "small" (288-transistor) circuits. Larger speedups are expected with larger circuits that yield longer vectors. A preliminary version of the revised 17000-statement program is expected to be released this summer.

C. COUPLING ACTIVITIES

1. SEMINARS

A review of our vector processing research was presented at the AFFDL.

2. CONSULTING

(a) Visiting scientist, AFFDL, to vectorized an explicit Navier-Stokes codes on the CYBER 205 and to study any associated I/O problems (-9/30/82).

(b) Visiting scientist, LANL, on vectorized Monte Carlo (5/1/81-4/30/82).

(c) Industrial consultant, Mobil Research and Development, on supercomputer procurement evaluation (5/1/81-1/15/82).

(d) Industrial consultant, Chevron Oil Field Research Co., on organization of vectorized sparse matrix algorithms (2/82).

3. OTHER

(a) A one-week short course at the University was presented in August, 1981, on High Speed Computation.

(b) Evaluation of proposals for the NASA Numerical Aerodynamic Simulator (NAS) was initiated, as an appointed member of a Technical Review Board (4/1/82)

D. REFERENCES

1. GRANT SUPPORTED PUBLICATIONS (5/1/81-4/30/82)

Journal Articles

[1] Calahan, D. A., "Performance of Linear Algebra Codes on the CRAY-1," SPE Journal, October, 1981, pp. 558-564.

Conference Publications

[2] Calahan, D. A, "Vectorized Direct Solvers for 2-D Grids," Proc. 6th Symposium on Reservoir Simulation, New Orleans, Feb. 1-2, 1982, pp. 489-506.

[3] Arya, S., and D. A. Calahan, "Optimal Scheduling of Assembly Language Kernels for Vector Processors," Proc. 19th Allerton Conference on Information, Control, and Computers, University of Illinois, October 1-2, 1981.

[4] Calahan, D. A., "A Research Summary: Six Years with the CRAY-1," Cray Scientists Meeting, (10th Anniversary, Cray Research, Inc.), April 5-7, Minneapolis.

Reports

[5] Calahan, D. A., "High Performance Banded and Profile Equation Solvers for the CRAY-1: I. The Unsymmetric Case," Report #160, Systems Engineering Laboratory, University of Michigan, February, 1982.

Other

[6] Calahan, D. A., "Direct Solution of Linear Equations on the CRAY-1," CRAY Channels (a production of Cray Research, Inc.), vol. 3, no. 1, July, 1981, pp. 105, 6.

2. OTHER REPORT REFERENCES

[7] Calahan, D. A., "Multi-level Vectorized Sparse Solution of LSI Circuits," Proc. IEEE Conf. on Circuits and Computers, Rye, N.Y., October 3, 1980, pp. 976-979.

[8] Calahan, D. A., "A Block-Oriented Sparse Equation Solver for the CRAY-1," Report #136, Systems Engineering Laboratory, University of Michigan, December, 1980.

[9] Duff, I. S., and J. H. Reid, "Experience of Sparse Matrix Codes on the CRAY-1," Report CSS 116, AERE Harwell, October, 1981.

[10] Calahan, D. A., P. G. Buning, and W. N. Joy, "Vectorized General Sparsity Algorithms with Backing Store," Report #96, Systems Engineering Laboratory, University of Michigan, January, 1977.

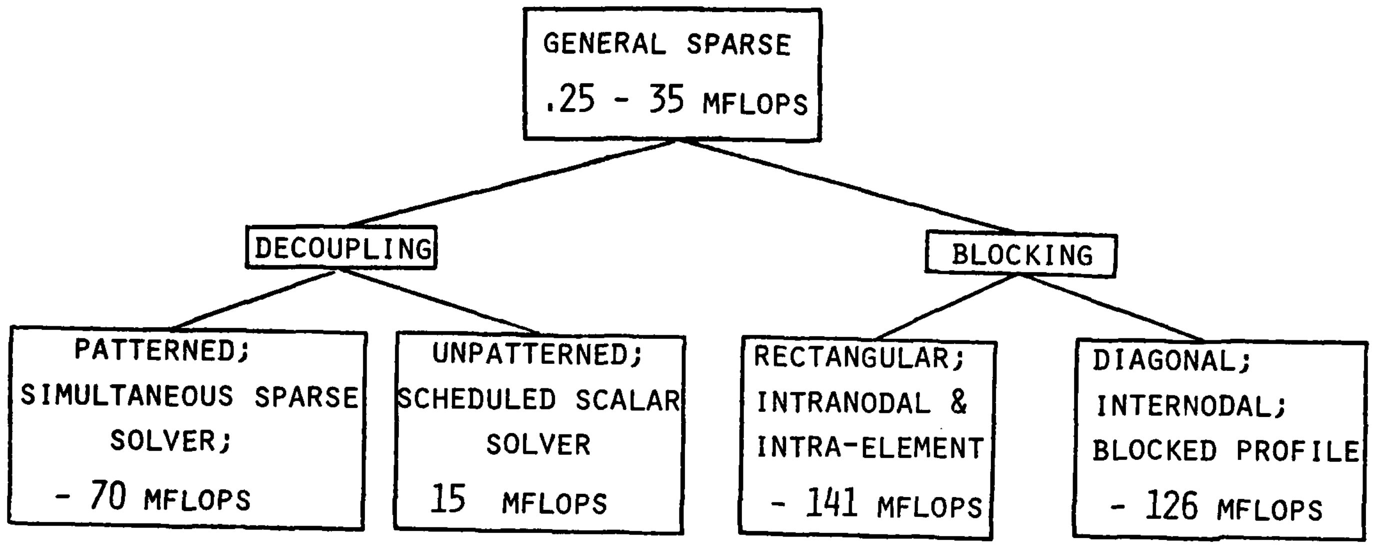
[11] Vladimirescu, A., and D. O. Pederson, "A Computer Program for the Analysis of LSI Circuits," Proceedings IEEE Intl. Symp. on Circuits and Systems, Chicago, Ill., April, 1981.

[12] Gustavson, F. G., W. M. Liniger, and R. A. Willoughby, "Symbolic Generation of an Optimal Crout Algorithm in Sparse Systems of Linear Equations," J. ACM, vol. 17, pp. 87-109.

[13] Jordan, T. L., and R. Fong, "Some Linear Algebra Codes and Their Performance on the CRAY-1," Report LA-6774, Los Alamos National Laboratory, June, 1977.

Figure 1

HIERARCHY AND CLASSIFICATION OF SPARSE MATRIX SOFTWARE



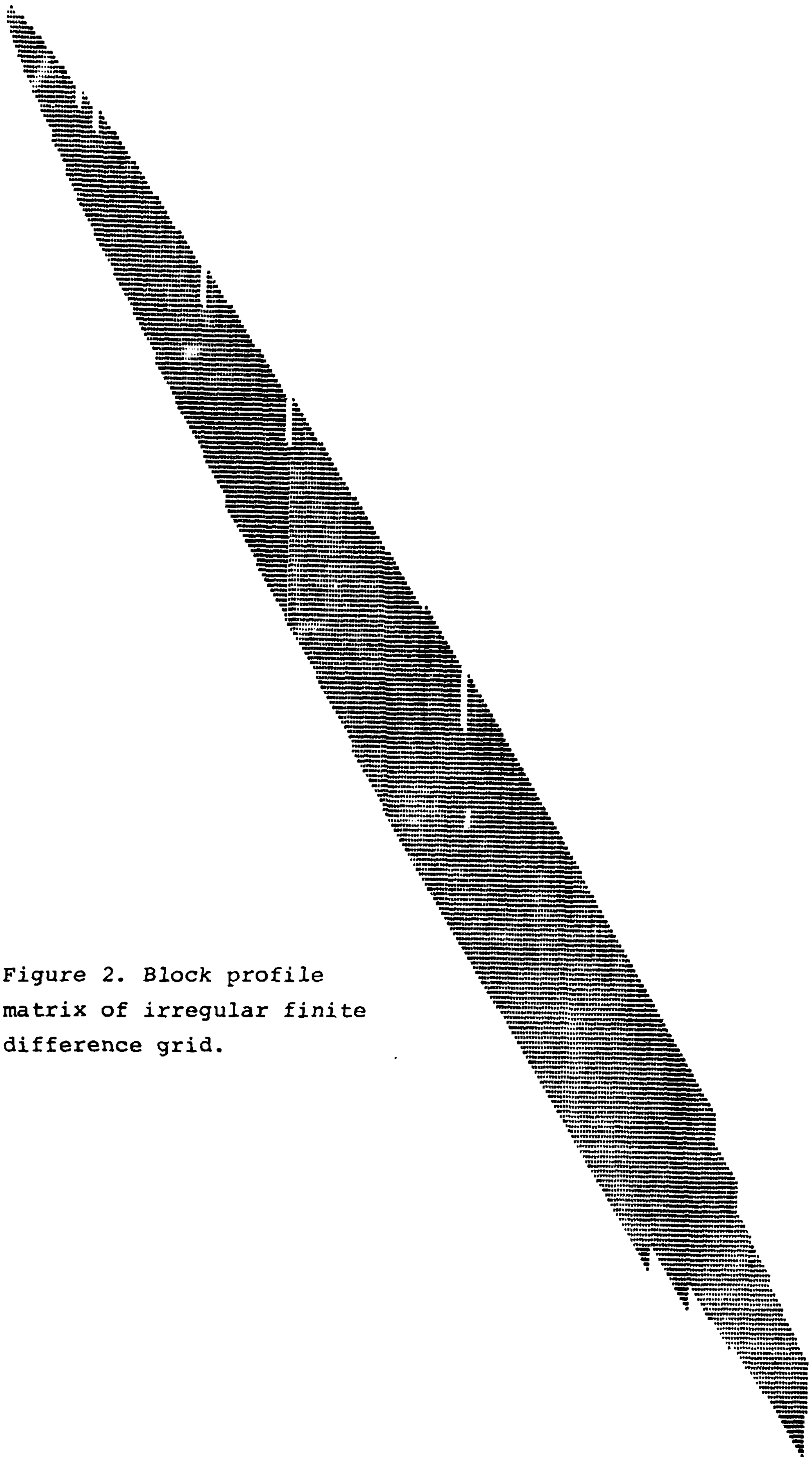


Figure 2. Block profile matrix of irregular finite difference grid.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 82-0553	2. GOVT ACCESSION NO. AD-A117112	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) VECTORIZED SPARSE ELIMINATION		5. TYPE OF REPORT & PERIOD COVERED INTERIM, 1 MAY 81-30 APR 82
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) D.A. Calahan		8. CONTRACT OR GRANT NUMBER(s) AFOSR-80-0158
9. PERFORMING ORGANIZATION NAME AND ADDRESS Systems Engineering Laboratory University of Michigan Ann Arbor MI 48109		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS PE61102F; 2304/A3
11. CONTROLLING OFFICE NAME AND ADDRESS Mathematical & Information Sciences Directorate Air Force Office of Scientific Research Bolling AFB DC 20332		12. REPORT DATE MAY 82
		13. NUMBER OF PAGES 9
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Figure 1 indicates how the single topic of general sparse matrix solution using scalar processors may be broken into specialized areas of study when implementation on vector architectures is considered.

First, highly sparse matrices, usually representing ODE/algebraic-modeled systems, are easily decoupled by re-ordering. At a minimum, locally-decoupled equations may be solved in pipelined scalar mode (see below); if the decoupled subsystems can be arranged (a) to have identical sparsity, and (CONTINUED)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ITEM #20, CONTINUED: (b) to be stored a constant stride apart, then a simultaneous sparse solver [7] may be invoked and a vector solution obtained.

As sparse systems become locally coupled - as occurs in finite element and finite difference problems - then vectors are easily defined within the coupled subsystems. It is worth making a further distinction between:

(a) intra-nodal or intra-element coupling, where the dimension of dense submatrices is proportional to the number of unknowns/node or unknowns/finite element, and

(b) inter-nodal or inter-element, where the coupling between grid nodes or finite elements determines the vector length.

Banded and profile matrices result from the latter. The associated vector lengths are the <u>products</u> of the number of unknown/node (element) and the number of coupled nodes. These lengths are therefore always longer than in the former case, so that common bandsolvers offer the highest performance of any sparse solvers.

In previous research, algorithms and CRAY-1 software have been developed on this grant for (Figure 1)

(a) general sparse matrices [10],

(b) patterned sparse matrices, in conjunction with a vectorized electronic circuit analysis program [7][11], and

(c) blocked matrices arising from intra-nodal coupling [B].

In addition, Duff and Reid [9] have converted a Fortran frontal ("intra-element") solver to the CRAY-1.

B

ILMED

8